

Some extensions of linear approximation and prediction problems for stationary processes

Ildar Ibragimov* Zakhar Kabluchko[†] Mikhail Lifshits[‡]

October 18, 2016

Abstract

Let $(B(t))_{t \in \Theta}$ with $\Theta = \mathbb{Z}$ or $\Theta = \mathbb{R}$ be a wide sense stationary process with discrete or continuous time. The classical linear prediction problem consists of finding an element in $\overline{\text{span}}\{B(s), s \leq t\}$ providing the best possible mean square approximation to the variable $B(\tau)$ with $\tau > t$.

In this article we investigate this and some other similar problems where, in addition to prediction quality, optimization takes into account other features of the objects we search for. One of the most motivating examples of this kind is an approximation of a stationary process B by a stationary differentiable process X taking into account the kinetic energy that X spends in its approximation efforts.

2010 AMS Mathematics Subject Classification: Primary: 60G10; Secondary: 60G15, 49J40, 41A00.

Key words and phrases: energy saving approximation, interpolation, prediction, wide sense stationary process.

*St.Petersburg Department of V.A.Steklov Institute of Mathematics of Russian Academy of Science, Fontanka embankment, 27, St.Petersburg, 191023, Russia, and St.Petersburg State University, 7/9 Universitetskaya nab., St.Petersburg, 199034 Russia email ibr32@pdmi.ras.ru

[†]Münster University, Orléans-Ring 10, 48149 Münster, Germany, email zakhar.kabluchko@uni-muenster.de

[‡]St.Petersburg State University, 7/9 Universitetskaya nab., St.Petersburg, 199034 Russia, email mikhail@lifshits.org and MAI, Linköping University.

1 Introduction and problem setting

1.1 Motivating example

Let $(B(t))_{t \in \Theta}$ with $\Theta = \mathbb{Z}$ or $\Theta = \mathbb{R}$ be a wide sense stationary process with discrete or continuous time. The classical linear prediction problem consists of finding an element in $\overline{\text{span}}\{B(s), s \leq t\}$ providing the best possible mean square approximation to the variable $B(\tau)$ with $\tau > t$, see [10] and [4, 5, 6, 13, 15, 16].

Below we investigate this and some other similar problems where, in addition to prediction quality, optimization takes into account other features of the objects we search for, such as the smoothness properties of approximation processes.

Here and elsewhere in the article $\overline{\text{span}}\{\cdot\}$ stands for the closed linear span of a set in a Hilbert space. All mentioned processes are assumed to be complex valued and all Hilbert spaces are assumed to be complex.

Example 1.1 (Approximation saving kinetic energy, [9]). By the instant *kinetic energy* of a process $(X(t))_{t \in \mathbb{R}}$ we understand just its squared derivative $|X'(t)|^2$. It is more than natural to search for an approximation of a given stationary process $(B(t))_{t \in \mathbb{R}}$ by a differentiable stationary process $(X(t))_{t \in \mathbb{R}}$ taking into account the kinetic energy that X spends in its approximation efforts. The goals of the approximation quality and energy saving may be naturally combined with averaging in time by minimization of the functional

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_0^N [|X(t) - B(t)|^2 + \alpha^2 |X'(t)|^2] dt.$$

Here $\alpha > 0$ is a fixed scaling regularization parameter balancing the quality of approximation and the spent energy.

If, additionally, the process $X(t) - B(t)$ and the derivative $X'(t)$ are stationary processes in the strict sense, in many situations ergodic theorem applies and the limit above is equal to $\mathbb{E} |X(0) - B(0)|^2 + \alpha^2 \mathbb{E} |X'(0)|^2$.

Therefore, we may simplify our task to solving the problem

$$\mathbb{E} |X(0) - B(0)|^2 + \alpha^2 \mathbb{E} |X'(0)|^2 \rightarrow \min, \quad (1)$$

and setting aside ergodicity issues.

The problem (1) makes sense either in a simpler *linear non-adaptive setting*, i.e. with

$$X(t) \in \overline{\text{span}}\{B(s), s \in \mathbb{R}\}, \quad t \in \mathbb{R},$$

or in *linear adaptive setting* by requiring additionally

$$X(t) \in \overline{\text{span}}\{B(s), s \leq t\}, \quad t \in \mathbb{R}.$$

In other words, this means that we only allow approximations based on the current and past values of B .

Let us start with a basic notation. Let $(B(t))_{t \in \Theta}$ be a complex-valued random process satisfying $\mathbb{E} B(t) = 0$, $\mathbb{E} |B(t)|^2 < \infty$ for all $t \in \Theta$.

Consider $H := \overline{\text{span}}\{B(t), t \in \Theta\}$ as a Hilbert space equipped with the scalar product $(\xi, \eta) = \mathbb{E}(\xi \overline{\eta})$. For $T \subset \Theta$ let $H(T) := \overline{\text{span}}\{B(t), t \in T\}$.

Furthermore, let L be a linear operator with values in H and defined on a linear subspace $\mathcal{D}(L) \subset H$. For a fixed $\tau \in \Theta$, consider the extremal problem

$$\mathbb{E} |Y - B(\tau)|^2 + \mathbb{E} |L(Y)|^2 \rightarrow \min, \quad (2)$$

where the minimum is taken over all $Y \in H(T) \cap \mathcal{D}(L)$. The first term in the sum describes approximation, prediction, or interpolation quality while the second term stands for additional properties of the object we are searching for, e.g. for the smoothness of the approximating process.

This is the most general form of the problem we are interested in. Below we specify the class of the considered processes to one-parametric wide sense stationary processes with discrete or continuous time, introduce appropriate class of operators L and explain in Subsection 1.5 why Example 1.1 is a special case of problem (2).

1.2 Spectral representation: brief reminder

Let now $(B(t))_{t \in \Theta}$ be the main object of our investigation – a centered wide sense stationary random process with univariate discrete ($\Theta = \mathbb{Z}$) or continuous ($\Theta = \mathbb{R}$) time. In case of continuous time we additionally assume that B is mean square continuous.

Here and elsewhere we assume that all random variables under consideration are centered. In particular, $\mathbb{E} B(t) = 0$ for all $t \in \Theta$.

By Khinchin theorem, the covariance function

$$K(t) := \mathbb{E} B(t) \overline{B(0)}$$

admits the spectral representation

$$K(t) := \int e^{itu} \mu(du).$$

Here and in the sequel integration in similar integrals is performed over the interval $[-\pi, \pi)$ in case of the discrete time processes and over real line \mathbb{R} in case of the continuous time processes. The finite measure μ on $[-\pi, \pi)$ or on \mathbb{R} , respectively, is the spectral measure of the process B . The process B itself admits the spectral representation as a stochastic integral

$$B(t) = \int e^{itu} \mathcal{W}(du) \quad (3)$$

where \mathcal{W} is an orthogonal random measure on $[-\pi, \pi)$ or on \mathbb{R} , respectively, with $\mathbb{E} |\mathcal{W}(A)|^2 = \mu(A)$.

Let

$$\mathcal{L} = L_2(\mu) = \left\{ \phi : \int |\phi(u)|^2 \mu(du) < \infty \right\}$$

be equipped with the usual scalar product

$$(\phi, \psi)_{\mathcal{L}} = \int \phi(u) \overline{\psi(u)} \mu(du).$$

Recall that for any

$$\xi = \int \phi(u) \mathcal{W}(du), \quad \eta = \int \psi(u) \mathcal{W}(du)$$

it is true that

$$\mathbb{E} \xi \overline{\eta} = \int \phi(u) \overline{\psi(u)} \mu(du) = (\phi, \psi)_{\mathcal{L}}.$$

It follows that the correspondence $B(t) \rightleftharpoons e^{itu}$ extends to the linear isometry between H and the closed linear span of the exponents in \mathcal{L} . Actually, the latter span coincides with entire space \mathcal{L} , cf. [6, Section 1.9], and we obtain a linear isometry between H and \mathcal{L} provided by stochastic integral. In other words, every element ξ of Hilbert space H can be represented as a stochastic integral

$$\xi = \int \phi_{\xi}(u) \mathcal{W}(du) \tag{4}$$

with some complex valued function $\phi_{\xi} \in \mathcal{L}$, and every random variable ξ admitting the representation (4) belongs to H .

An analogous theory exists for processes with wide sense stationary increments. Let B be such process with zero mean (for continuous time we additionally assume mean square continuity). Similarly to (3), the process B admits a spectral representation

$$B(t) = B_0 + B_1 t + \int (e^{itu} - 1) \mathcal{W}(du)$$

where $\mathcal{W}(du)$ is an orthogonal random measure controlled by the spectral measure μ and B_0, B_1 are centered random variables uncorrelated with \mathcal{W} , see [15, p.213]. Notice that in case of processes with wide sense stationary increments the spectral measure μ need not be finite but it must satisfy $\mu\{0\} = 0$ and Lévy's integrability condition

$$\int_{-\pi}^{\pi} u^2 \mu(du) < \infty$$

for discrete time and

$$\int_{\mathbb{R}} \min\{u^2, 1\} \mu(du) < \infty$$

for continuous time.

In the following we let $B_0 = B_1 = 0$ because for prediction problems we handle here the finite rank part is uninteresting. We also do not lose any interesting example with this restriction. Therefore, we consider the processes

$$B(t) = \int (e^{itu} - 1) \mathcal{W}(du).$$

1.3 Probabilistic problem setting

The operators L we are going to handle are those of the form

$$L\xi = L \left(\int \phi_\xi(u) \mathcal{W}(du) \right) := \int \ell(u) \phi_\xi(u) \mathcal{W}(du), \quad (5)$$

where ℓ is a measurable function on \mathbb{R} or on $[-\pi, \pi)$, respectively. The domain $\mathcal{D}(L)$ consists of ξ such that

$$\mathbb{E} |L\xi|^2 = \|L\xi\|_H^2 = \int |\ell(u) \phi_\xi(u)|^2 \mu(du) < \infty.$$

Such operators are often called *linear filters* while the function ℓ is called the *frequency characteristic* of a filter.

Below we consider problem (2) applied to wide sense stationary processes with discrete or continuous time and operators L from (5). For the space $H(T)$ we consider a variety of choices. Most typically, we take $H(T) = H := \overline{\text{span}}\{B(s), -\infty < s < \infty\}$, the space generated by all variables, or $H(T) = H_t := H((-\infty, t]) = \overline{\text{span}}\{B(s), s \leq t\}$, the space generated by the past of the process, or $H(T) = H_t^\circ := \overline{\text{span}}\{B(s), |s| \geq t\}$.

In problem (2), we take the value of B at some point τ as a subject of approximation. When B is a wide sense stationary process, we may take $\tau = 0$ without loss of generality.

Therefore, three following variations of problem (2) are considered below.

Problem I (approximation):

$$\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \rightarrow \min, \quad Y \in H.$$

Problem II (prediction):

$$\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \rightarrow \min, \quad Y \in H_t.$$

Problem III (interpolation):

$$\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \rightarrow \min, \quad Y \in H_t^\circ.$$

Notice that, due to the presence of L , Problems II and III represent an extension of the classical prediction and interpolation problems. As for Problem I, once L is omitted, it is trivial. In our setting it is also easy but provides non-trivial results (even in the simplest cases) and therefore is sufficiently interesting.

Sometimes we call the setting of Problem II adaptive, because the best approximation is based on (adapted to) the known past values of the process. Opposite to this, the setting of Problem I is called non-adaptive.

In the classical case, i.e. with $L = 0$, Problem II, as stated here, is non-trivial only for negative t . However, in presence of L it makes sense for arbitrary t .

1.4 Analytic problem setting

Due to the spectral representation (4) problems I – III admit the following analytic setting.

Problem I':

$$\int |\psi(u) - 1|^2 \mu(du) + \int |\ell(u)\psi(u)|^2 \mu(du) \rightarrow \min, \quad \psi \in \mathcal{L}. \quad (6)$$

Problem II':

$$\int |\psi(u) - 1|^2 \mu(du) + \int |\ell(u)\psi(u)|^2 \mu(du) \rightarrow \min, \quad \psi \in \overline{\text{span}}\{e^{isu}, s \leq t\}. \quad (7)$$

Problem III':

$$\int |\psi(u) - 1|^2 \mu(du) + \int |\ell(u)\psi(u)|^2 \mu(du) \rightarrow \min, \quad \psi \in \overline{\text{span}}\{e^{isu}, |s| \geq t\}. \quad (8)$$

The spans in Problems II' and III' are taken in \mathcal{L} .

1.5 Energy saving approximation as a special case of extended prediction problem

Consider the setting of Example 1.1: given a zero mean wide sense stationary process $B = (B(t))_{t \in \mathbb{R}}$ with spectral representation (3), the problem is to minimize the functional

$$\mathbb{E} |X(0) - B(0)|^2 + \alpha^2 \mathbb{E} |X'(0)|^2$$

over all mean square differentiable processes $X = (X(t))_{t \in \mathbb{R}}$ such that the processes X and B are jointly wide sense stationary. The latter means that each of them is wide sense stationary and also the cross-covariance $\mathbb{E} X(t)\overline{B(s)}$ depends only on $t - s$.

First of all, we show that while solving this minimization problem one may only consider approximating processes of special type, namely,

$$\tilde{X}(t) = \int e^{itu} \psi(u) \mathcal{W}(du), \quad \psi \in \mathcal{L}. \quad (9)$$

Indeed, for arbitrary X , we may decompose its initial value as $X(0) = X^\perp + \tilde{X}(0)$ with $\tilde{X}(0) \in H$, X^\perp orthogonal to H . By representation (4) there exists $\psi \in \mathcal{L}$ such that

$$\tilde{X}(0) = \int \psi(u) \mathcal{W}(du).$$

For this ψ define the process $\tilde{X}(t)$ by (9). We show that the process \tilde{X} is at least as good as X in the sense of (1).

Due to the joint wide sense stationarity, for any s, t we have

$$\begin{aligned} \mathbb{E} X(t) \overline{B(s)} &= \mathbb{E} X(0) \overline{B(s-t)} = \mathbb{E} \tilde{X}(0) \overline{B(s-t)} = \int \psi(u) e^{i(t-s)u} \mu(du); \\ \mathbb{E} \tilde{X}(t) \overline{B(s)} &= \int [e^{itu} \psi(u)] e^{-isu} \mu(du) = \int \psi(u) e^{i(t-s)u} \mu(du). \end{aligned}$$

It follows that $X(t) - \tilde{X}(t)$ is orthogonal to $B(s)$ for each s , hence, it is orthogonal to H . Furthermore, it is easy to show that if X is mean square differentiable then so are its components \tilde{X} and $X - \tilde{X}$. For their derivatives, we know that

$$\tilde{X}'(t) = \int e^{itu} (iu) \psi(u) \mathcal{W}(du) \in H$$

and $(X - \tilde{X})'(t)$ is orthogonal to H . Hence,

$$\begin{aligned} &\mathbb{E} |X(0) - B(0)|^2 + \alpha^2 \mathbb{E} |X'(0)|^2 \\ &= \mathbb{E} |(X(0) - \tilde{X}(0)) + (\tilde{X}(0) - B(0))|^2 + \alpha^2 \mathbb{E} |(X'(0) - \tilde{X}'(0)) + \tilde{X}'(0)|^2 \\ &= \mathbb{E} |X(0) - \tilde{X}(0)|^2 + \mathbb{E} |\tilde{X}(0) - B(0)|^2 \\ &\quad + \alpha^2 \mathbb{E} |(X' - \tilde{X}')(0)|^2 + \alpha^2 \mathbb{E} |\tilde{X}'(0)|^2 \\ &\geq \mathbb{E} |\tilde{X}(0) - B(0)|^2 + \alpha^2 \mathbb{E} |\tilde{X}'(0)|^2. \end{aligned}$$

Therefore, \tilde{X} is at least as good for (1), as X .

Finally, notice that for the processes defined by (9) the expression in (1) is equal to

$$\int |\psi(u) - 1|^2 \mu(du) + \int |\ell(u) \psi(u)|^2 \mu(du)$$

with $\ell(u) = \alpha i u$, exactly as in the analytical versions of our problems (6) and (7). In the non-adaptive version of the approximation problem we have to optimize over \mathcal{L} , as in (6), while for adaptive version the requirement $\tilde{X} \in H_t$ for all t is satisfied iff $\psi \in \overline{\text{span}}\{e^{isu}, s \leq 0\}$, as in (7).

One may consider other types of energy, e.g. based on higher order derivatives of X . This option leads to the same problems with arbitrary polynomials ℓ .

For discrete time case it is natural to replace the derivative X' by the difference $X(1) - X(0)$. Then we obtain the same problem with $\ell(u) = \alpha(e^{iu} - 1)$, for $u \in [-\pi, \pi)$.

Examples of optimal non-adaptive energy saving approximation are given in Section 4 below.

One may also consider the energy saving approximation for the processes with wide sense stationary increments. Consider such a process B and its approximation X such that $(X(t), B(t))_{t \in \Theta}$ with $\Theta = \mathbb{Z}$ or $\Theta = \mathbb{R}$ is a two-dimensional process with wide sense stationary increments and $(X(t) - B(t))_{t \in \Theta}$ is a wide sense stationary process. Since $B(0) = 0$, the analogue of (1) is

$$\mathbb{E} |X(0)|^2 + \alpha^2 \mathbb{E} |X'(0)|^2 \rightarrow \min. \quad (10)$$

Similarly to the case of wide sense stationary processes one can show that, analogously to (9), it is sufficient only to consider approximating processes of the special form

$$\tilde{X}(t) := \int (e^{itu}\psi(u) - 1) \mathcal{W}(du), \quad \psi - 1 \in \mathcal{L}.$$

Then problem (10) takes familiar analytical form

$$\int |\psi(u) - 1|^2 \mu(du) + \int |\psi(u)(iu)|^2 \mu(du) \rightarrow \min$$

with requirements $\psi - 1 \in \mathcal{L}$ for non-adaptive setting and

$$e^{itu}\psi(u) - 1 \in \overline{\text{span}}\{e^{isu} - 1, s \leq t\}$$

in adaptive setting. The latter may be simplified to

$$\psi - 1 \in \overline{\text{span}}\{e^{isu} - 1, s \leq 0\}.$$

2 Abstract Hilbert space setting

The basic matters about our problems such as the existence of the solution or its uniqueness are easier to handle in a more abstract setting. A formal extension of problem (2) looks as follows. Let H be a separable Hilbert space with the corresponding scalar product (\cdot, \cdot) and norm $\|\cdot\|$. Let L

be a linear operator taking values in H and defined on a linear subspace $\mathcal{D}(L) \subset H$. Consider a problem

$$G(y) := \|y - x\|^2 + \|Ly\|^2 \rightarrow \min. \quad (11)$$

Here x is a given element of H and minimum is taken over all $y \in H_0 \cap \mathcal{D}(L)$ where $H_0 \subset H$ is a given closed linear subspace.

The following results are probably well known, yet for completeness we give their proofs in Section 7.

Proposition 2.1 *If L is a closed operator then the problem (11) has a solution $\xi \in H_0 \cap \mathcal{D}(L)$.*

Proposition 2.2 *The problem (11) has at most one solution.*

Remark 2.3 Unlike Proposition 2.1, the assertion of Proposition 2.2 holds without additional assumptions on the operator L .

Proposition 2.4 *Assume that in problem (11) we have $H_0 = H$ and L is a closed operator with the domain dense in H . Then the unique solution of (11) exists and is given by the formula*

$$\xi = (I + L^*L)^{-1}x,$$

where $I : H \rightarrow H$ is the identity operator.

Proposition 2.5 *If ξ is a solution of problem (11), then ξ provides the unique solution of equations*

$$(\xi - x, h) + (L\xi, Lh) = 0 \quad \text{for all } h \in H_0 \cap \mathcal{D}(L). \quad (12)$$

Remark 2.6 If the operator L is bounded, one may rewrite equations (12) as

$$((I + L^*L)\xi - x, h) = 0 \quad \text{for all } h \in H_0,$$

where I is the identity operator in H .

3 Solution of the non-adaptive problem

Theorem 3.1 *Let B be a centered wide sense stationary process with discrete or continuous time. Let L be a linear filter (5) with arbitrary measurable frequency characteristic $\ell(\cdot)$. Then the unique solution of Problem I exists and is given by the formula*

$$\xi = \int \frac{1}{1 + |\ell(u)|^2} \mathcal{W}(du). \quad (13)$$

The error of optimal approximation, i.e. the minimum in Problem I, and in its equivalent form (6), is given by

$$\begin{aligned}\sigma^2 &:= \mathbb{E} |\xi - B(0)|^2 + \mathbb{E} |L\xi|^2 = \int \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} \mu(du) \\ &= \mathbb{E} |B(0)|^2 - \int \frac{\mu(du)}{1 + |\ell(u)|^2}.\end{aligned}\quad (14)$$

Proof: The operators of type (5) are clearly closed and have a dense domain. Therefore, Proposition 2.1 provides the existence of solution. Furthermore, Proposition 2.2 confirms that the solution is unique. Proposition 2.4 states the form of the solution

$$\xi = (I + L^*L)^{-1}B(0).$$

By using the definition of L , for any $Y \in H$ we easily obtain

$$(I + L^*L)^{-1}Y = \int \frac{1}{1 + |\ell(u)|^2} \phi_Y(u) \mathcal{W}(du).$$

For $Y = B(0)$ we have $\phi_Y(u) \equiv 1$, thus (13) follows. Finally, by isometric property,

$$\begin{aligned}\sigma^2 &= \mathbb{E} |\xi - B(0)|^2 + \mathbb{E} |L\xi|^2 \\ &= \int \left[\left| \frac{1}{1 + |\ell(u)|^2} - 1 \right|^2 + \left| \frac{\ell(u)}{1 + |\ell(u)|^2} \right|^2 \right] \mu(du) \\ &= \int \frac{|\ell(u)|^4 + |\ell(u)|^2}{(1 + |\ell(u)|^2)^2} \mu(du) = \int \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} \mu(du),\end{aligned}$$

as claimed in (14). \square

Remark 3.2 For equivalent Problem I' one can arrive to the same conclusion as in Theorem 3.1 in a fairly elementary way. Using the full square identity

$$|\psi(u) - 1|^2 + |\ell(u)\psi(u)|^2 = (1 + |\ell(u)|^2) \left| \psi(u) - \frac{1}{1 + |\ell(u)|^2} \right|^2 + \frac{|\ell(u)|^2}{1 + |\ell(u)|^2},$$

one immediately observes that

$$\psi_\xi(u) := \frac{1}{1 + |\ell(u)|^2} \quad (15)$$

solves Problem I', while the error is given by (14).

Remark 3.3 For processes with wide sense stationary increments the extremal problem is the same, hence the solution (13) is the same. It should be noticed however that the solution is correct only if the quantity σ^2 above is finite (for finite measure μ it is always finite but for infinite measure and some choices of ℓ it may be infinite). For example if ℓ is a polynomial without free term, $\ell(u) = \sum_{k=1}^m c_k u^k$, then σ^2 above is finite. This includes kinetic energy case $\ell(u) = i\alpha u$. Otherwise, if

$$\int \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} \mu(du) = \infty,$$

the quantity in Problem I' is infinite for all admissible ψ .

4 Some examples of non-adaptive approximation

In this section we illustrate general results by some typical examples. In all examples we consider kinetic energy, i.e. we let $\ell(u) = \alpha(e^{iu} - 1)$ in the discrete time case and $\ell(u) = i\alpha u$ in the continuous time.

For discrete time we get

$$\begin{aligned} |\ell(u)|^2 + 1 &= \alpha^2(e^{iu} - 1)(e^{-iu} - 1) + 1 \\ &= \frac{\alpha^2}{\beta}(e^{iu} - \beta)(e^{-iu} - \beta) \end{aligned}$$

where

$$\beta = \frac{2\alpha^2 + 1 + \sqrt{1 + 4\alpha^2}}{2\alpha^2} > 1 \quad (16)$$

is the larger root of the equation

$$\beta^2 - \frac{2\alpha^2 + 1}{\alpha^2} \beta + 1 = 0.$$

For the integrand in the solution (13) of the non-adaptive problem, we easily derive an expansion

$$\begin{aligned} \frac{1}{|\ell(u)|^2 + 1} &= \frac{\beta}{\alpha^2} \frac{1}{(e^{iu} - \beta)(e^{-iu} - \beta)} \\ &= \frac{1}{\sqrt{1 + 4\alpha^2}} \left(1 + \sum_{k=1}^{\infty} \beta^{-k} (e^{iku} + e^{-iku}) \right). \end{aligned} \quad (17)$$

By plugging this expression into (13), it follows that the solution of discrete non-adaptive problem involving kinetic energy is given by the moving average with bilateral geometric progression weight:

$$\xi = \frac{1}{\sqrt{1 + 4\alpha^2}} \left(B(0) + \sum_{k=1}^{\infty} \beta^{-k} (B(k) + B(-k)) \right). \quad (18)$$

By (14), the error of optimal non-adaptive approximation in the discrete time case is given by

$$\sigma^2 = \int_{-\pi}^{\pi} \frac{\alpha^2 |e^{iu} - 1|^2}{\alpha^2 |e^{iu} - 1|^2 + 1} \mu(du). \quad (19)$$

For continuous time we get similar results. By using inverse Fourier transform, we have

$$\frac{1}{|\ell(u)|^2 + 1} = \frac{1}{\alpha^2 u^2 + 1} = \frac{1}{2\alpha} \int_{\mathbb{R}} \exp\{i\tau u - |\tau|/\alpha\} d\tau.$$

By plugging this expression into (13), it follows that the solution of continuous non-adaptive problem involving kinetic energy is given by the moving average

$$\xi = \frac{1}{2\alpha} \int_{\mathbb{R}} \exp\{-|\tau|/\alpha\} B(\tau) d\tau. \quad (20)$$

By (14), the error of optimal non-adaptive approximation in the continuous time case is given by

$$\sigma^2 = \int_{\mathbb{R}} \frac{\alpha^2 u^2}{\alpha^2 u^2 + 1} \mu(du). \quad (21)$$

Notice that both solutions (18) and (20) are indeed non-adaptive at all because they involve future values of B . Let us also stress that these solutions formulae are the same for any spectral (covariance) structure of B . The formulae (20) and (21) were obtained earlier in [9].

We start with discrete time examples.

Example 4.1 A sequence $(B(t))_{t \in \mathbb{Z}}$ of centered non-correlated random variables with constant variance $V \geq 0$ has the spectral measure

$$\mu(du) := \frac{V du}{2\pi}.$$

Surprisingly, the answer to the non-adaptive problem taking kinetic energy into account even for this sequence is already non-trivial, since the best non-adaptive approximation is given by the series (18). The error formula (19) yields

$$\sigma^2 = V \int_{-\pi}^{\pi} \frac{\alpha^2 |e^{iu} - 1|^2 du}{2\pi \alpha^2 |e^{iu} - 1|^2 + 1} = V \left(1 - \frac{1}{\sqrt{1 + 4\alpha^2}} \right),$$

cf. an extension below in (24).

Example 4.2 A sequence of random variables $(B(t))_{t \in \mathbb{Z}}$ is called autoregressive, if it satisfies the equation $B(t) = \rho B(t-1) + \xi(t)$, where $|\rho| < 1$ and $(\xi(t))_{t \in \mathbb{Z}}$ is a sequence of centered non-correlated random variables with some variance V . In this case we have a representation

$$B(t) = \sum_{j=0}^{\infty} \rho^j \xi(t-j), \quad t \in \mathbb{Z}.$$

Given the spectral representation $\xi(t) = \int_{-\pi}^{\pi} e^{itu} \mathcal{W}(du)$ from the previous example, we obtain

$$B(t) = \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \rho^j e^{i(t-j)u} \mathcal{W}(du) = \int_{-\pi}^{\pi} \frac{1}{1 - \rho e^{-iu}} e^{itu} \mathcal{W}(du).$$

We see that the spectral measure for B is

$$\mu(du) := \frac{V du}{2\pi |1 - \rho e^{-iu}|^2}. \quad (22)$$

The best non-adaptive approximation is given by the series (18). By (19) and (22), the error of non-adaptive approximation is

$$\begin{aligned} \sigma^2 &= \frac{V}{2\pi} \int_{-\pi}^{\pi} \frac{\alpha^2 |e^{iu} - 1|^2}{\alpha^2 |e^{iu} - 1|^2 + 1} \frac{du}{|1 - \rho e^{-iu}|^2} \\ &= \frac{V}{2\pi} \left[\int_{-\pi}^{\pi} \frac{du}{|1 - \rho e^{-iu}|^2} - \int_{-\pi}^{\pi} \frac{1}{\alpha^2 |e^{iu} - 1|^2 + 1} \frac{du}{|1 - \rho e^{-iu}|^2} \right]. \end{aligned}$$

By using the expansion

$$\frac{1}{|1 - \rho e^{-iu}|^2} = \frac{1}{1 - \rho^2} \left(1 + \sum_{k=1}^{\infty} \rho^k (e^{iku} + e^{-iku}) \right) \quad (23)$$

we obtain immediately that

$$\int_{-\pi}^{\pi} \frac{du}{|1 - \rho e^{-iu}|^2} = \frac{2\pi}{1 - \rho^2}.$$

Moreover, it follows from (17) and (23) that

$$\int_{-\pi}^{\pi} \frac{1}{\alpha^2 |e^{iu} - 1|^2 + 1} \frac{du}{|1 - \rho e^{-iu}|^2} = \frac{2\pi}{\sqrt{1 + 4\alpha^2}(1 - \rho^2)} \frac{\beta + \rho}{\beta - \rho}$$

with $\beta = \beta(\alpha)$ defined in (16). Finally,

$$\sigma^2 = \frac{V}{1 - \rho^2} \left(1 - \frac{1}{\sqrt{1 + 4\alpha^2}} \frac{\beta + \rho}{\beta - \rho} \right). \quad (24)$$

Notice that Example 4.1 is a special case of this one with $\rho = 0$.

Example 4.3 We call a sequence of random variables $(B(t))_{t \in \mathbb{Z}}$ a simplest moving average sequence if it admits a representation $B(t) = \xi(t) + \rho \xi(t-1)$ where $\xi(t)$ is the same as in Example 4.2. Proceeding as above, we obtain

$$B(t) = \int_{-\pi}^{\pi} (1 + \rho e^{-iu}) e^{itu} \mathcal{W}(du), \quad t \in \mathbb{Z}.$$

We see that the spectral measure for B is

$$\mu(du) := \frac{V|1 + \rho e^{-iu}|^2 du}{2\pi}. \quad (25)$$

The best non-adaptive approximation is given by (18). By (25) and (19), the error of non-adaptive approximation is

$$\begin{aligned} \sigma^2 &= \frac{V}{2\pi} \int_{-\pi}^{\pi} \frac{\alpha^2 |e^{iu} - 1|^2}{\alpha^2 |e^{iu} - 1|^2 + 1} |1 + \rho e^{-iu}|^2 du \\ &= \frac{V}{2\pi} \int_{-\pi}^{\pi} \left[1 - \frac{1}{\alpha^2 |e^{iu} - 1|^2 + 1} \right] |1 + \rho e^{-iu}|^2 du \\ &= \frac{V}{2\pi} \left[\int_{-\pi}^{\pi} |1 + \rho e^{-iu}|^2 du - \int_{-\pi}^{\pi} \frac{|1 + \rho e^{-iu}|^2}{\alpha^2 |e^{iu} - 1|^2 + 1} du \right]. \end{aligned}$$

We easily get

$$\int_{-\pi}^{\pi} |1 + \rho e^{-iu}|^2 du = \int_{-\pi}^{\pi} (1 + \rho^2 + \rho(e^{-iu} + e^{iu})) du = 2\pi(1 + \rho^2)$$

and, by using (17),

$$\begin{aligned} &\int_{-\pi}^{\pi} \frac{|1 + \rho e^{-iu}|^2}{\alpha^2 |e^{iu} - 1|^2 + 1} du = \frac{1}{\sqrt{1 + 4\alpha^2}} \times \\ &\times \int_{-\pi}^{\pi} (1 + \rho^2 + \rho(e^{-iu} + e^{iu})) \left(1 + \sum_{k=1}^{\infty} \beta^{-k} (e^{iku} + e^{-iku}) \right) du \\ &= \frac{2\pi}{\sqrt{1 + 4\alpha^2}} \left(1 + \rho^2 + \frac{2\rho}{\beta} \right), \end{aligned}$$

whereas

$$\sigma^2 = V \left(1 + \rho^2 - \frac{1}{\sqrt{1 + 4\alpha^2}} \left(1 + \rho^2 + \frac{2\rho}{\beta} \right) \right)$$

with $\beta = \beta(\alpha)$ defined in (16).

Example 4.4 Consider the partial sums of a sequence of centered non-correlated random variables $(\xi(j))_{j \geq 1}$ each having variance V . Let $B(0) = 0$ and $B(t) := \sum_{j=1}^t \xi(j)$, $t \in \mathbb{N}$. Given the spectral representation in the form $\xi_j = \int_{-\pi}^{\pi} e^{iju} \mathcal{W}(du)$, we have

$$B(t) = \int_{-\pi}^{\pi} \left(\sum_{j=1}^t e^{iju} \right) \mathcal{W}(du) = \int_{-\pi}^{\pi} \frac{e^{itu} - 1}{e^{iu} - 1} (e^{itu} - 1) \mathcal{W}(du)$$

and we obtain the spectral measure

$$\mu(du) := \frac{V du}{2\pi |e^{iu} - 1|^2}. \quad (26)$$

The best non-adaptive approximation is given by (18). By (26) and (19), the corresponding approximation error of

$$\sigma^2 = \frac{V}{2\pi} \int_{-\pi}^{\pi} \frac{\alpha^2 du}{\alpha^2 |e^{iu} - 1|^2 + 1}.$$

Furthermore, by using expansion (17) we have

$$\sigma^2 = \frac{V\alpha^2}{2\pi} \cdot \frac{2\pi}{\sqrt{1+4\alpha^2}} = \frac{V\alpha^2}{\sqrt{1+4\alpha^2}}. \quad (27)$$

We pass now to continuous time examples.

Example 4.5 The Ornstein–Uhlenbeck process is a centered Gaussian stationary process with covariance $K_B(t) = e^{-|t|/2}$ and the spectral measure

$$\mu(du) := \frac{2du}{\pi(4u^2 + 1)}. \quad (28)$$

Since we are developing only a linear theory, we do not need Gaussianity assumption and may call Ornstein–Uhlenbeck any wide sense stationary process having the mentioned covariance and spectral measure.

The best non-adaptive approximation is given by (20). By (28) and (21), the error of non-adaptive approximation is

$$\sigma^2 = \int_{\mathbb{R}} \frac{\alpha^2 u^2}{\alpha^2 u^2 + 1} \frac{2du}{\pi(4u^2 + 1)} = \frac{\alpha}{2 + \alpha}.$$

Example 4.6 Fractional Brownian motion $(B^H(t))_{t \in \mathbb{R}}$, $0 < H \leq 1$, is a centered Gaussian process with covariance

$$\text{Cov}(B^H(t_1), B^H(t_2)) = \frac{1}{2} (|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}).$$

For any process with this covariance (interesting non-Gaussian examples of this type are also known, see e.g. Telecom processes in [8, 11]) the spectral measure is

$$\mu(du) := \frac{M_H du}{|u|^{2H+1}} \quad (29)$$

where $M_H = \frac{\Gamma(2H+1) \sin(\pi H)}{2\pi}$.

The best non-adaptive approximation is given by (20). By (29) and (21), the error of non-adaptive approximation is

$$\begin{aligned}
\sigma^2 &= \int_{\mathbb{R}} \frac{\alpha^2 u^2}{\alpha^2 u^2 + 1} \frac{M_H du}{|u|^{2H+1}} = M_H \alpha^2 \int_{\mathbb{R}} \frac{|u|^{1-2H} du}{\alpha^2 u^2 + 1} \\
&= 2M_H \alpha^{2H} \int_0^\infty \frac{w^{1-2H} dw}{w^2 + 1} = M_H \alpha^{2H} \int_0^\infty \frac{v^{-H} dv}{v + 1} \\
&= M_H \alpha^{2H} \cdot \frac{\pi}{\sin(\pi H)} = \frac{\Gamma(2H + 1) \alpha^{2H}}{2}.
\end{aligned} \tag{30}$$

This result was obtained in [9].

Example 4.7 Consider a centered Lévy process $(B(t))_{t \geq 0}$ with finite variance (this class includes Wiener process and centered Poisson processes of any constant intensity). Let $\text{Var } B(1) = V$. For any such process the spectral measure is

$$\mu(du) := \frac{V du}{2\pi u^2}. \tag{31}$$

This is a continuous version of Example 4.4, as well as a special case of Example 4.6 with $H = \frac{1}{2}$. Notice, however, that in the more delicate problems of *adaptive* approximation (that are not studied here) the cases $H = \frac{1}{2}$ and $H \neq \frac{1}{2}$ are totally different.

The best non-adaptive approximation is given by (20). The calculation of non-adaptive approximation error based on (31) and (21) is a special case of (30) with $H = \frac{1}{2}$, up to a scaling constant V . We have thus

$$\sigma^2 = \frac{\alpha V}{2}. \tag{32}$$

Finally, notice that there is a natural interplay between continuous time and discrete time approximation. Let $(B(t))_{t \in \mathbb{R}}$ be a continuous time wide sense stationary process. For any small $\delta > 0$ consider its discrete time version $(B_\delta(s))_{s \in \mathbb{Z}} := (B(\delta s))_{s \in \mathbb{Z}}$. The discrete time counterpart for kinetic energy $\alpha^2 X'(t)^2$ of approximating process is $\frac{\alpha^2 (X(\delta(s+1)) - X(\delta s))^2}{\delta^2}$, thus one should consider discrete time approximation with parameter $\alpha_\delta := \frac{\alpha}{\delta}$. Using (16), we see that

$$\beta_\delta := \beta(\alpha_\delta) = 1 + \frac{1 + o(1)}{\alpha_\delta}, \quad \text{as } \alpha_\delta \rightarrow \infty,$$

hence $\beta_\delta^{\alpha_\delta} \rightarrow e$ as $\alpha_\delta \rightarrow \infty$.

Therefore, for the best discrete time approximations (18) of B_δ we have

$$\begin{aligned} X_\delta(s) &= \frac{1}{\sqrt{1+4\alpha_\delta^2}} \left(B(s) + \sum_{k=1}^{\infty} \beta_\delta^{-k} (B(s+k\delta) + B(s-k\delta)) \right) \\ &= \frac{(1+o(1))\delta}{2\alpha} \left(B_\delta(s) + \sum_{k=1}^{\infty} [\beta_\delta^{\alpha_\delta}]^{-k\delta/\alpha} (B(s+k\delta) + B(s-k\delta)) \right) \\ &\rightarrow \frac{1}{2\alpha} \int_{\mathbb{R}} \exp\{-|\tau|/\alpha\} B(s+\tau) d\tau, \quad \text{as } \delta \rightarrow 0, \end{aligned}$$

which is the solution of continuous time approximation problem (20). Similarly, one has the convergence of the optimal errors of approximation, cf. e.g. (27) and (32).

5 An extension of A.N. Kolmogorov and M.G. Krein theorems on error-free prediction

5.1 Discrete time

In this subsection we assume that $(B(t))_{t \in \mathbb{Z}}$ is a wide sense centered stationary sequence and μ is its spectral measure. Let us represent μ as the sum $\mu = \mu_a + \mu_s$ of its absolutely continuous and singular components. We denote by f_a the density of μ_a with respect to the Lebesgue measure.

Consider Problem II and let

$$\sigma^2(t) := \inf_{Y \in H_t} \{ \mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \}, \quad t \in \mathbb{Z},$$

be the corresponding prediction errors. We also let $\sigma^2(\infty)$ denote the similar quantity with H_t replaced by H . It is easy to see that the sequence $\sigma^2(t)$ is non-increasing in t and

$$\lim_{t \rightarrow +\infty} \sigma^2(t) = \sigma^2(\infty).$$

For the classical prediction problem, i.e. for $L = 0$, by Kolmogorov's theorem (singularity criterion, see [13, Chapter II, Theorem 5.4]) we have

$$\sigma^2(t) = \sigma^2(\infty) = 0 \quad \text{for all } t \in \mathbb{Z},$$

iff

$$\int_{-\pi}^{\pi} |\ln f_a(u)| du = \infty. \quad (33)$$

In our case, for $L \neq 0$, we have $\sigma^2(t) \geq \sigma^2(\infty) > 0$ unless $\ell(\cdot) \equiv 0$ μ -a.s. Therefore, we state the problem as follows: when $\sigma^2(t) = \sigma^2(\infty)$ holds for a given $t \in \mathbb{Z}$? In other words: when approximation based on the knowledge of the process up to time t works as well as the one based on the knowledge of the whole process?

Theorem 5.1 *If (33) holds, then we have $\sigma^2(t) = \sigma^2(\infty)$ for all $t \in \mathbb{Z}$.*

Proof: If (33) holds, then for all t we have $H_t = H$, see e.g. [4, Chapter XII, Section 4] or [13, Chapter II, Section 2]. Therefore, by Theorem 3.1

$$\sigma^2(t) = \sigma^2(\infty) = \int_{-\pi}^{\pi} \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} \mu(du).$$

□

Theorem 5.2 *If the process B is such that the density f_a satisfies*

$$\int_{-\pi}^{\pi} |\ln f_a(u)| du < \infty, \quad (34)$$

then for every fixed $t \in \mathbb{Z}$ we have $\sigma^2(t) = \sigma^2(\infty)$ iff the function $\frac{1}{1+|\ell(u)|^2}$ is a trigonometric polynomial of degree not exceeding t , i.e.

$$\frac{1}{1 + |\ell(u)|^2} = \sum_{|j| \leq t} b_j e^{iju}$$

Lebesgue-a.e. with some coefficients $b_j \in \mathbb{C}$.

In particular, if $t < 0$, then $\sigma^2(t) < \sigma^2(\infty)$; equality $\sigma^2(0) = \sigma^2(\infty)$ holds iff $|\ell(\cdot)|$ is a constant Lebesgue-a.e.

Proof: The analytic form for prediction error is

$$\sigma^2(t) = \inf_{\psi \in \mathcal{L}(t)} \int_{-\pi}^{\pi} \{ |\psi(u) - 1|^2 + |\ell(u)\psi(u)|^2 \} \mu(du),$$

where $\mathcal{L}(t) = \overline{\text{span}}\{e^{isu}, s \leq t\}$ in \mathcal{L} .

The solution $\psi_t(u)$ of our problem is unique by Proposition 2.2; we know from (15) that for $t = \infty$ the solution is $\psi_\infty(u) = \frac{1}{1+|\ell(u)|^2}$. It follows that $\sigma^2(t) = \sigma^2(\infty)$ iff $\psi_t = \psi_\infty$, i.e. iff $\psi_\infty \in \mathcal{L}(t)$. The latter is equivalent to the existence of the trigonometric polynomials

$$\vartheta_k(u) := \sum_{j \leq t} a_{j,k} e^{iju}$$

such that

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} \left| \vartheta_k(u) - \frac{1}{1 + |\ell(u)|^2} \right|^2 \mu(du) = 0.$$

It follows that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} \left| \overline{\vartheta_k(u)} - \frac{1}{1 + |\ell(u)|^2} \right|^2 f_a(u) du \\ &= \lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} \left| \vartheta_k(u) - \frac{1}{1 + |\ell(u)|^2} \right|^2 f_a(u) du = 0. \end{aligned} \quad (35)$$

Due to assumption (34) the density f_a admits a representation

$$f_a(u) = |g_*(e^{iu})|^2$$

where $g_*(e^{iu})$, $u \in [-\pi, \pi)$, is the boundary value of the function

$$g(z) := \exp \left\{ \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln f_a(u) \frac{e^{iu} + z}{e^{iu} - z} du \right\}, \quad |z| < 1,$$

which is an analytic function in the unit disc $\mathbb{D} := \{z \in \mathbb{C}, |z| < 1\}$. In other words, g is an outer function from the Hardy class $\mathcal{H}_2(\mathbb{D})$; we refer to [14, Chapter 17] for the facts and definitions mentioned in this subsection concerning $\mathcal{H}_2(\mathbb{D})$, outer and inner functions. Notice that $\frac{1}{g}$ also is an outer analytic function.

Rewrite (35) as

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} \left| \overline{\vartheta_k(u)} g_*(e^{iu}) - \frac{g_*(e^{iu})}{1 + |\ell(u)|^2} \right|^2 du = 0.$$

Assume for a while that $t \leq 0$. Then $\overline{\vartheta_k} g_*$ also is the boundary value of a function from $\mathcal{H}_2(\mathbb{D})$. Since the class of such boundary functions is closed in L_2 (with respect to Lebesgue measure), this implies that

$$h_*(e^{iu}) := \frac{g_*(e^{iu})}{1 + |\ell(u)|^2}$$

is the boundary value of a function $h \in \mathcal{H}_2(\mathbb{D})$. Moreover, we have a power series representation

$$h(z) = \sum_{j \geq -t} h_j z^j, \quad z \in \mathbb{D}. \quad (36)$$

Let us denote

$$A_1(u) := \frac{1}{1 + |\ell(u)|^2} = h_*(e^{iu}) \cdot \frac{1}{g_*(e^{iu})}, \quad u \in [-\pi, \pi).$$

The function $e^{iu} \mapsto A_1(u)$ admits an analytic continuation from the unit circle to \mathbb{D} given by $A(z) = h(z) \cdot \frac{1}{g(z)}$. Notice that in the power series representation

$$A(z) = \sum_{j=0}^{\infty} a_j z^j, \quad z \in \mathbb{D},$$

the terms with $j < -t$ vanish due to (36). We have therefore

$$A(z) = \sum_{j \geq -t}^{\infty} a_j z^j, \quad z \in \mathbb{D}.$$

Let us prove that $A(\cdot)$ is bounded on \mathbb{D} . Write the factorization $h = M_h \cdot Q_h$ where M_h is an inner function and Q_h is an outer function. Then $A = M_h \cdot \frac{Q_h}{g}$. The function M_h is bounded on \mathbb{D} by the definition of an inner function while $\frac{Q_h}{g}$ is an outer function with bounded boundary values, because Lebesgue-a.e.

$$\left| \left(\frac{Q_h}{g} \right)_* (e^{iu}) \right| = \left| \frac{A_1(u)}{(M_h)_*(e^{iu})} \right| = |A_1(u)| \leq 1.$$

Since for outer functions the boundedness on the boundary implies, via Poisson kernel representation, the boundedness on \mathbb{D} , we see that the factor $\frac{Q_h}{g}$ is also bounded on \mathbb{D} . We conclude that $A(\cdot)$ is bounded on \mathbb{D} .

For each $r \in (0, 1)$ consider the function

$$A_r(u) := A(re^{iu}) = \sum_{j \geq -t}^{\infty} a_j r^j e^{iju}, \quad u \in [-\pi, \pi).$$

Since A is bounded, the family $\{A_r\}_{0 < r < 1}$ is uniformly bounded. Since $A_r \rightarrow A_1$ Lebesgue-a.e., as $r \nearrow 1$, the convergence also holds in L_2 . In particular, all Fourier coefficients converge and we have

$$A_1(u) = \frac{1}{1 + |\ell(u)|^2} = \sum_{j \geq -t} a_j e^{iju}.$$

Since the left hand side is real, for $t < 0$ the latter representation is impossible. For $t = 0$ it is only possible when both sides are equal (Lebesgue-a.e.) to the constant a_0 .

For $t > 0$ the same reasonings give a representation

$$\frac{e^{itu}}{1 + |\ell(u)|^2} = \sum_{j \geq 0} a_j e^{iju}$$

which implies that

$$\frac{1}{1 + |\ell(u)|^2} = \sum_{j=-t}^t a_{j+t} e^{iju}$$

is a trigonometric polynomial of degree not exceeding t .

The converse assertion is obvious: if for $t \geq 0$ we have a representation

$$\psi_\infty(u) = \frac{1}{1 + |\ell(u)|^2} = \sum_{j=-t}^t b_j e^{iju},$$

then $\psi_\infty \in \mathcal{L}(t)$ by the definition of $\mathcal{L}(t)$. □

Theorems 5.1 and 5.2 immediately yield the following final result.

Theorem 5.3 *Let B be a discrete time, wide sense stationary process. Let L be a linear filter with frequency characteristic $\ell(\cdot)$. Then for every fixed $t \in \mathbb{Z}$ the equality $\sigma^2(t) = \sigma^2(\infty)$ holds iff either (33) holds, or (34) holds and $\frac{1}{1+|\ell(u)|^2}$ is a trigonometric polynomial of degree not exceeding t .*

5.2 Continuous time

In this subsection we assume that $(B(t))_{t \in \mathbb{R}}$ is a continuous time, mean square continuous, wide sense stationary process and μ is its spectral measure. As before, we represent μ as the sum $\mu = \mu_a + \mu_s$ of its absolutely continuous and singular components, denote f_a the density of μ_a with respect to Lebesgue measure and let

$$\sigma^2(t) := \inf_{Y \in H_t} \{ \mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \}, \quad t \in \mathbb{R},$$

denote the corresponding prediction errors. We also let $\sigma^2(\infty)$ denote the similar quantity with H_t replaced by H .

The statement analogous to Theorem 5.3 is as follows.

Theorem 5.4 *Let B be a continuous time, mean square continuous, wide sense stationary process. Let L be a linear filter with frequency characteristic $\ell(\cdot)$. Then for every fixed $t \in \mathbb{R}$ the equality*

$$\sigma^2(t) = \sigma^2(\infty) = \int \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} \mu(du)$$

holds iff either

(a)

$$\int_{-\infty}^{\infty} \frac{|\ln f_a(u)|}{1 + u^2} du = \infty \tag{37}$$

holds or

(b)

$$\int_{-\infty}^{\infty} \frac{|\ln f_a(u)|}{1 + u^2} du < \infty \tag{38}$$

holds, $t > 0$ and $\frac{1}{1+|\ell(u)|^2}$ is a restriction (to \mathbb{R}) of an entire analytic function of exponential type not exceeding t , or

(c) *inequality (38) holds, $t = 0$, and $|\ell(\cdot)|$ is Lebesgue a.e. equal to a constant.*

Proof: If (37) holds, then by M.G. Krein singularity criterion, we have $H_t = H$ for all $t \in \mathbb{R}$, see e.g. [4, Chapter XII, Section 4] or [13, Chapter II, Section 2], and the assertion a) of the theorem follows.

Let $\Pi := \{z \in \mathbb{C} : \Im(z) > 0\}$ denote the upper half-plane. If (38) holds, then we have a representation $f_a(u) = |g_*(u)|^2$, where $g_*(u)$ is the boundary value of the function

$$g(z) := \exp \left\{ \frac{1}{2\pi i} \int_{-\infty}^{\infty} \ln f_a(u) \frac{1+zu}{z-u} \frac{du}{1+u^2} \right\}.$$

Therefore g is an outer function from Hardy class $\mathcal{H}_2(\Pi)$; we refer to [7, Chapter 8] for the facts and definitions mentioned in this subsection concerning Hardy classes on Π and related outer and inner functions.

Let $t \leq 0$. The same arguments as those given in the proof of Theorem 5.2 show that

$$h_*(u) := \frac{g_*(u)}{1 + |\ell(u)|^2}$$

is the boundary value of a function $h \in \mathcal{H}_2(\Pi)$.

The function

$$A_*(u) := \frac{1}{1 + |\ell(u)|^2} = h_*(u) \cdot \frac{1}{g_*(u)}, \quad u \in \mathbb{R}, \quad (39)$$

admits an analytic continuation from the real line to Π given by $A(z) = h(z) \cdot \frac{1}{g(z)}$.

Let us prove that $A(\cdot)$ is bounded on Π . Write the factorization $h = M_h \cdot Q_h$ where M_h is an inner function and Q_h is an outer function. Then $A = M_h \cdot \frac{Q_h}{g}$. The function M_h is bounded on Π by the definition of an inner function while $\frac{Q_h}{g}$ is an outer function with bounded boundary values, because Lebesgue-a.e.

$$\left| \left(\frac{Q_h}{g} \right)_*(u) \right| = \left| \frac{A_*(u)}{(M_h)_*(u)} \right| = |A_*(u)| \leq 1, \quad u \in \mathbb{R}.$$

Since for outer functions the boundedness on the boundary implies, via Poisson kernel representation, the boundedness on Π , we see that the factor $\frac{Q_h}{g}$ is also bounded on Π . We conclude that $A(\cdot)$ is bounded on Π . In other words, $A \in \mathcal{H}_\infty(\Pi)$.

Furthermore, the function A admits an analytic reflection to the lower half-plane $\Pi_- := \{z \in \mathbb{C} : \bar{z} \in \Pi\}$ by letting $A_-(z) := \overline{A(\bar{z})}$. This reflection agrees with A on the real line because the boundary values $A_*(u), u \in \mathbb{R}$, are real.

Consider now an auxiliary function $\mathcal{S}_*(u) := \frac{\sin u}{u} e^{iu}$ on \mathbb{R} and its analytic continuation $\mathcal{S}(z) := \frac{\sin z}{z} e^{iz} \in \mathcal{H}_2(\Pi)$. Then $\mathcal{S} \cdot A \in \mathcal{H}_2(\Pi)$ and the corresponding boundary function $\mathcal{S}_* \cdot A_* \in L_2(\mathbb{R})$. According to the Fourier representation of the elements of $\mathcal{H}_2(\Pi)$ and that of their boundary values

we have

$$\begin{aligned} \mathcal{S}_*(u)A_*(u) &= \frac{\sin u}{u} e^{iu} A_*(u) \\ &= \int_0^\infty e^{ixu} q(x) dx, \quad u \in \mathbb{R}, \text{ Lebesgue-a.e.}, \end{aligned} \quad (40)$$

$$\begin{aligned} \mathcal{S}(z)A(z) &= \frac{\sin z}{z} e^{iz} A(z) \\ &= \int_0^\infty e^{ixz} q(x) dx, \quad z \in \Pi, \end{aligned} \quad (41)$$

with some $q(\cdot) \in L_2(\mathbb{R}_+)$. We may rewrite (40) as

$$\frac{\sin u}{u} A_*(u) = \int_{-1}^\infty e^{iyu} q(y+1) dy.$$

Moreover, since the left hand side is real, its Fourier transform is symmetric. Therefore, $q(\cdot+1)$ must vanish on $[1, \infty)$, i.e. $q(\cdot)$ must vanish on $[2, \infty)$. Thus (41) writes as

$$\mathcal{S}(z) A(z) = \int_0^2 e^{ixz} q(x) dx := Q(z) \quad (42)$$

and Q is an entire function.

Consider the holomorphic function

$$V(z) := \frac{Q(z)}{\mathcal{S}(z)} = \frac{ze^{-iz}Q(z)}{\sin z}.$$

Since $V = A$ on Π and $V = A_*$ on \mathbb{R} , we see that V is bounded on $\Pi \cup \mathbb{R}$.

Proceeding in the same way with the lower half-plane Π_- instead of Π , we find "another" holomorphic function V_- such that $V_- = A_-$ on Π_- and $V_- = A_* = V$ on \mathbb{R} . The latter equality yields $V = V_-$ on \mathbb{C} ; moreover, V is bounded on \mathbb{C} . Hence V is a constant and $A_* = V$ is a constant, too, as required in the assertion c) of our theorem.

If (38) holds and $t > 0$, the same reasoning leads to a representation analogue to (39), namely,

$$A_*(u) := \frac{1}{1 + |\ell(u)|^2} = e^{-itu} \mathcal{A}_*(u), \quad u \in \mathbb{R},$$

where \mathcal{A}_* is the boundary function of some $\mathcal{A} \in \mathcal{H}_\infty(\Pi)$.

Using that $\mathcal{S} \cdot \mathcal{A} \in \mathcal{H}_2(\Pi)$ and proceeding as before, we have a representation analogue to (42)

$$\mathcal{S}(z) \mathcal{A}(z) = \int_0^{2(t+1)} e^{izx} q(x) dx := Q(z)$$

and Q is an entire function. Let

$$V(z) = \frac{Q(z)e^{-itz}}{\mathcal{S}(z)} = \mathcal{A}(z)e^{-itz}.$$

Then $V = A_*$ on \mathbb{R} and we have

$$|V(z)| \leq e^{t|\Im(z)|} \|A\|_\infty, \quad z \in \Pi.$$

Proceeding in the same way with the lower half-plane Π_- instead of Π , we find "another" holomorphic function V_- such that

$$|V_-(z)| \leq e^{t|\Im(z)|} \|A_-\|_\infty, \quad z \in \Pi_-,$$

and $V_- = A_* = V$ on \mathbb{R} . The latter equality yields $V = V_-$ on \mathbb{C} . We conclude that

$$|V(z)| \leq e^{t|\Im(z)|} \|A\|_\infty, \quad z \in \mathbb{C},$$

i.e. V is an entire function of exponential type not exceeding t , as required in the assertion b) of our theorem.

If (38) holds and $t < 0$, we still see from the previous reasoning that ψ_∞ must be a constant. Due to M.G. Krein's regularity criterion (see [13, Chapter III, Theorem 2.4]), we know that under (38) constants do not belong to $\mathcal{L}(t)$ with $t < 0$. Hence the equality $\sigma^2(t) = \sigma^2(\infty)$ is not possible for $t < 0$.

Now we prove the sufficiency. Assume that $t > 0$ and that $A_*(u) := \frac{1}{1+|\ell(u)|^2}$ is a restriction (to \mathbb{R}) of an entire analytic function $A(\cdot)$ of exponential type not exceeding t .

Write

$$A_*(u) = A_*(0) + u \cdot \frac{A_*(u) - A_*(0)}{u} := A_*(0) + u \cdot \widetilde{A}_*(u).$$

It is sufficient to show that the function $u \mapsto u \cdot \widetilde{A}_*(u)$ belongs to H_t . Furthermore, since we have

$$\frac{1 - e^{-i\delta u}}{i\delta} \widetilde{A}_*(u) \rightarrow u \cdot \widetilde{A}_*(u), \quad \text{as } \delta \rightarrow 0,$$

in $L_2(\mathbb{R}, \mu)$, it is sufficient to show that $\widetilde{A}_* \in H_t$.

Notice that \widetilde{A}_* belongs to $L_2(\mathbb{R}^1)$ w.r.t. Lebesgue measure and is a restriction of the analytic function of exponential type not exceeding t given by $\widehat{A}(z) := \frac{A(z) - A(0)}{z}$. Hence, by Paley–Wiener theorem (see [1, Chapter IV]) we have a representation

$$\widetilde{A}_*(u) = \widehat{A}(u) = \int_{-t}^t a(\tau) e^{i\tau u} d\tau$$

with $a(\cdot) \in L_2[-t, t] \subset \widehat{L_1}[-t, t]$. Since exponentials $u \mapsto e^{i\tau u}$ belong to H_t as $\tau \leq t$, it follows that $\widehat{A}_* \in H_t$.

This concludes the proof of our theorem. \square

Remark 5.5 In 1923, S.N. Bernstein introduced a class of entire functions of exponential type not exceeding t and bounded on the real line, cf. [3] or [1, Chapter 4, Section 83]. The functions that appear in Theorem 5.4 belong to this class.

5.3 An extension of regularity criterion

Consider again the discrete time case. We handle wide sense stationary sequences and use the notation from Subsection 5.1. Let $\sigma_0^2(t)$ be the prediction error for $B(0)$ given the past H_t in the classical prediction problem (with $L = 0$). The sequence B is called *regular*, if we have

$$\lim_{t \rightarrow -\infty} \sigma_0^2(t) = \mathbb{E} |B(0)|^2.$$

By the classical Kolmogorov regularity criterion, see [13, Chapter II, Theorem 5.1], a wide sense stationary sequence B is regular iff its spectral measure is absolutely continuous and (34) is verified.

The following result provides an extension of this assertion to our settings.

Theorem 5.6 *Let B be a wide sense stationary sequence. Let L be a linear filter with frequency characteristic $\ell(\cdot)$. If (34) holds, then*

$$\lim_{t \rightarrow -\infty} \sigma^2(t) = \mathbb{E} |B(0)|^2 - \int_{-\pi}^{\pi} \frac{\mu_s(du)}{1 + |\ell(u)|^2}.$$

Proof: Consider first the sequences with absolutely continuous spectral measure. In that case, by Kolmogorov's criterion, B is regular. Hence,

$$\lim_{t \rightarrow -\infty} \sigma^2(t) \geq \lim_{t \rightarrow -\infty} \inf_{Y \in H_t} \mathbb{E} |Y - B(0)|^2 = \mathbb{E} |B(0)|^2.$$

On the other hand, since for each $t \in \mathbb{Z}$ it is true $Y = 0 \in H_t$, we have $\sigma^2(t) \leq \mathbb{E} |B(0)|^2$ and the theorem is proved in the form

$$\lim_{t \rightarrow -\infty} \sigma^2(t) = \mathbb{E} |B(0)|^2. \quad (43)$$

In the general case, by [13, Chapter II, Theorem 2.2] our sequence splits into a sum $B = B^{(a)} + B^{(s)}$ of mutually orthogonal wide sense stationary processes such that the regular part $B^{(a)}$ has the spectral measure μ_a , the singular part $B^{(s)}$ has the spectral measure μ_s and the corresponding spaces $H_t^{(a)} := \overline{\text{span}}\{B^{(a)}(v), v \leq t\}$, $H_t^{(s)} := \overline{\text{span}}\{B^{(s)}(v), v \leq t\}$ are not only

orthogonal but also satisfy subordination inclusions $H_t^{(a)} \subset H_t$, $H_t^{(s)} \subset H_t$. The latter yield $H_t = H_t^{(a)} \oplus H_t^{(s)}$.

Moreover, for any $\xi \in H_t$ representation (5) implies

$$\begin{aligned} \mathbb{E} |L\xi|^2 &= \int_{-\pi}^{\pi} |\ell(u)\phi_{\xi}(u)|^2 \mu(du) \\ &= \int_{-\pi}^{\pi} |\ell(u)\phi_{\xi}(u)|^2 \mu_a(du) + \int_{-\pi}^{\pi} |\ell(u)\phi_{\xi}(u)|^2 \mu_s(du) \\ &= \mathbb{E} |L\xi^{(a)}|^2 + \mathbb{E} |L\xi^{(s)}|^2, \end{aligned}$$

where $\xi^{(a)}, \xi^{(s)}$ denote the projections of ξ onto $H_t^{(a)}$ and $H_t^{(s)}$, respectively. Therefore, for any $\xi \in H_t$ we have

$$\begin{aligned} &\mathbb{E} |\xi - B(0)|^2 + \mathbb{E} |L\xi|^2 \\ &= \mathbb{E} |\xi^{(a)} - B^{(a)}(0)|^2 + \mathbb{E} |\xi^{(s)} - B^{(s)}(0)|^2 \\ &\quad + \mathbb{E} |L\xi^{(a)}|^2 + \mathbb{E} |L\xi^{(s)}|^2. \end{aligned}$$

In this situation, the minimization problem splits into two independent ones and is solved by $\xi_t = \xi_t^{(a)} + \xi_t^{(s)}$, where $\xi_t^{(a)}, \xi_t^{(s)}$ are the solutions for the processes $B^{(a)}, B^{(s)}$, respectively.

For the extended prediction errors we obtain, by using Theorem 5.1,

$$\sigma^2(t) = \sigma^{(a)2}(t) + \sigma^{(s)2}(t) = \sigma^{(a)2}(t) + \sigma^{(s)2}(\infty). \quad (44)$$

By applying consequently (44) and (43), we have

$$\begin{aligned} \lim_{t \rightarrow -\infty} \sigma^2(t) &= \lim_{t \rightarrow -\infty} \sigma^{(a)2}(t) + \sigma^{(s)2}(\infty) \\ &= \mathbb{E} |B^{(a)}(0)|^2 + \int_{-\pi}^{\pi} \frac{|\ell(u)|^2 \mu_s(du)}{1 + |\ell(u)|^2} \\ &= \mathbb{E} |B(0)|^2 - \mathbb{E} |B^{(s)}(0)|^2 + \int_{-\pi}^{\pi} \frac{|\ell(u)|^2 \mu_s(du)}{1 + |\ell(u)|^2} \\ &= \mathbb{E} |B(0)|^2 + \int_{-\pi}^{\pi} \left(\frac{|\ell(u)|^2}{1 + |\ell(u)|^2} - 1 \right) \mu_s(du) \\ &= \mathbb{E} |B(0)|^2 - \int_{-\pi}^{\pi} \frac{\mu_s(du)}{1 + |\ell(u)|^2}, \end{aligned}$$

as claimed. □

6 Interpolation

Consider the simplest case of interpolation problem (our Problem III) in discrete time. Let $(B(t))_{t \in \mathbb{Z}}$ be a wide sense stationary sequence having

spectral density f and let L be a linear filter with frequency characteristic $\ell(\cdot)$. Consider the extremal problem

$$\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2 \rightarrow \min, \quad Y \in H_1^\circ.$$

Recall that $H_1^\circ = \overline{\text{span}}\{B(s), |s| \geq 1\}$. Let

$$\sigma_{\text{int}}^2 = \inf_{Y \in H_1^\circ} (\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2)$$

denote the interpolation error.

The classical case of this problem, i.e. $L = 0$, was considered by A.N. Kolmogorov [10]. He proved that precise extrapolation with $\sigma_{\text{int}}^2 = 0$ is possible iff

$$\int_{-\pi}^{\pi} \frac{du}{f(u)} = \infty. \quad (45)$$

If the integral in (45) is convergent, then

$$\sigma_{\text{int}}^2 = 4\pi^2 \left(\int_{-\pi}^{\pi} \frac{du}{f(u)} \right)^{-1}.$$

We extend this result to the case of general L as follows.

Theorem 6.1 *If (45) holds, then*

$$\sigma_{\text{int}}^2 = \int_{-\pi}^{\pi} \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} f(u) du.$$

Otherwise,

$$\begin{aligned} \sigma_{\text{int}}^2 &= \int_{-\pi}^{\pi} \frac{|\ell(u)|^2 f(u) du}{1 + |\ell(u)|^2} \\ &+ \left(\int_{-\pi}^{\pi} \frac{du}{1 + |\ell(u)|^2} \right)^2 \left(\int_{-\pi}^{\pi} \frac{du}{f(u)(1 + |\ell(u)|^2)} \right)^{-1}. \end{aligned}$$

Proof: If (45) holds, then by Kolmogorov's theorem we have $B(0) \in H_1^\circ$, thus $H_1^\circ = H$ and by (14)

$$\begin{aligned} \sigma_{\text{int}}^2 &= \inf_{Y \in H} (\mathbb{E} |Y - B(0)|^2 + \mathbb{E} |LY|^2) \\ &= \int_{-\pi}^{\pi} \frac{|\ell(u)|^2}{1 + |\ell(u)|^2} f(u) du, \end{aligned}$$

proving the first assertion of the theorem.

Assume now that

$$\int_{-\pi}^{\pi} \frac{du}{f(u)} < \infty. \quad (46)$$

Let us define a function ϕ on $[-\pi, \pi)$ by the relations

$$\begin{aligned}\phi(u) &:= \frac{c + f(u)}{f(u)(1 + |\ell(u)|^2)}, \\ c &= - \left(\int_{-\pi}^{\pi} \frac{du}{1 + |\ell(u)|^2} \right) \left(\int_{-\pi}^{\pi} \frac{du}{f(u)(1 + |\ell(u)|^2)} \right)^{-1}.\end{aligned}$$

We will prove that the random variable

$$\xi = \int_{-\pi}^{\pi} \phi(u) \mathcal{W}(du)$$

solves our interpolation problem. To this aim, according to Propositions 2.1, 2.2, and 2.5, it is sufficient to prove that $\xi \in \mathcal{D}(L)$, that $\xi \in H_1^\circ$, and that ξ satisfies equations (12) of Proposition 2.5.

First, we have

$$\begin{aligned}\mathbb{E} |L\xi|^2 &= \int_{-\pi}^{\pi} |\phi(u)\ell(u)|^2 f(u) du \\ &= \int_{-\pi}^{\pi} \frac{(c + f(u))^2}{f(u)} \frac{|\ell(u)|^2}{(1 + |\ell(u)|^2)^2} du \\ &\leq \frac{1}{4} \int_{-\pi}^{\pi} \frac{(c + f(u))^2}{f(u)} du \\ &= \frac{c^2}{4} \int_{-\pi}^{\pi} \frac{du}{f(u)} + \pi c + \frac{1}{4} \int_{-\pi}^{\pi} f(u) du < \infty,\end{aligned}$$

whence $\xi \in \mathcal{D}(L)$.

Second, we show that $\xi \in H_1^\circ$. Consider an orthogonal decomposition $\xi = \eta + \eta^\perp$ with $\eta \in H_1^\circ$, $\eta^\perp \in (H_1^\circ)^\perp$. We also have the corresponding analytic decomposition $\phi = \psi + \psi^\perp$ with $\psi \in \mathcal{L}^\circ := \overline{\text{span}}\{e^{isu}, |s| \geq 1\}$ and $\psi^\perp \in (\mathcal{L}^\circ)^\perp$.

Let us show that any $h \in \mathcal{L}^\circ$ satisfies equation

$$\int_{-\pi}^{\pi} h(u) du = 0. \tag{47}$$

Indeed under assumption (46) the linear functional $h \mapsto \int_{-\pi}^{\pi} h(u) du$ is bounded and continuous on \mathcal{L} because by Hölder's inequality

$$\begin{aligned}\left| \int_{-\pi}^{\pi} h(u) du \right|^2 &\leq \int_{-\pi}^{\pi} \frac{du}{f(u)} \int_{-\pi}^{\pi} |h(u)|^2 f(u) du \\ &= \int_{-\pi}^{\pi} \frac{du}{f(u)} \|h\|_{\mathcal{L}}^2.\end{aligned}$$

Therefore, equality (47) which is true for every exponent from the set $\{h(u) = e^{isu}, |s| \geq 1\}$, extends to their span \mathcal{L}° . In particular, we obtain

$$\int_{-\pi}^{\pi} \psi(u) du = 0.$$

Since the constant c in the definition of ϕ was chosen so that

$$\int_{-\pi}^{\pi} \phi(u) du = 0, \quad (48)$$

we obtain

$$\int_{-\pi}^{\pi} \psi^{\perp}(u) du = \int_{-\pi}^{\pi} (\phi - \psi)(u) du = 0. \quad (49)$$

On the other hand, we have $\psi^{\perp} f \in L_1[-\pi, \pi]$ and $\psi^{\perp} \in (\mathcal{L}^{\circ})^{\perp}$. The latter means that

$$\int_{-\pi}^{\pi} \psi^{\perp}(u) e^{isu} f(u) du = 0, \quad s \in \mathbb{Z}, s \neq 0.$$

Therefore $\psi^{\perp} f$ is a constant, say, $\psi^{\perp} f = a$. By plugging $\psi^{\perp} = \frac{a}{f}$ into (49) we obtain $a = 0$. It follows that $\psi^{\perp} = 0$ and $\phi = \psi \in \mathcal{L}^{\circ}$, which is equivalent to $\xi \in H_1^{\circ}$.

It remains to check that ξ satisfies equations (12). The analytical form of these equations is

$$\int_{-\pi}^{\pi} [\phi(u) - 1 + |\ell(u)|^2 \phi(u)] h(u) f(u) du = 0, \quad h \in \mathcal{L}^{\circ}. \quad (50)$$

By the definition of ϕ we have

$$\phi(u) - 1 + |\ell(u)|^2 \phi(u) = \frac{c}{f(u)}.$$

Therefore, (50) reduces to

$$\int_{-\pi}^{\pi} h(u) du = 0, \quad h \in \mathcal{L}^{\circ}.$$

The latter was already verified in (47), and we have proved that ξ is a solution of interpolation problem. Now the direct computation using the definitions of ϕ and c , as well as (48), shows

$$\begin{aligned} \sigma_{\text{int}}^2 &= \int_{-\pi}^{\pi} [(\phi(u) - 1)^2 + |\ell(u)|^2 \phi(u)^2] f(u) du \\ &= \int_{-\pi}^{\pi} [\phi(u)^2 (1 + |\ell(u)|^2) + (1 - 2\phi(u))] f(u) du \\ &= \int_{-\pi}^{\pi} [\phi(u)(c + f(u)) + (1 - 2\phi(u))f(u)] du \\ &= \int_{-\pi}^{\pi} (1 - \phi(u))f(u) du = \int_{-\pi}^{\pi} \frac{|\ell(u)|^2 f(u) - c}{1 + |\ell(u)|^2} du \\ &= \int_{-\pi}^{\pi} \frac{|\ell(u)|^2 f(u)}{1 + |\ell(u)|^2} du \\ &\quad + \left(\int_{-\pi}^{\pi} \frac{du}{1 + |\ell(u)|^2} \right)^2 \left(\int_{-\pi}^{\pi} \frac{du}{f(u)(1 + |\ell(u)|^2)} \right)^{-1}, \end{aligned}$$

as claimed in the theorem's assertion. \square

Remark 6.2 The first term in σ_{int}^2 is just the optimal error (14) in the easier optimization problem with $Y \in H$. The second term is the price we must pay for optimization over the smaller space H_1° instead of H .

7 Proofs for abstract Hilbert space setting

Proof of Proposition 2.1: Not that the set $H_0 \cap \mathcal{D}(L)$ is non-empty, since it contains zero. Let

$$\sigma^2 := \inf_{y \in H_0 \cap \mathcal{D}(L)} G(y).$$

There exists a sequence $(\xi_n)_{n \in \mathbb{N}}$ in $H_0 \cap \mathcal{D}(L)$ such that

$$\lim_{n \rightarrow \infty} G(\xi_n) = \sigma^2.$$

Clearly, for all $m, n \in \mathbb{N}$ we have $(\xi_m + \xi_n)/2 \in H_0 \cap \mathcal{D}(L)$ and by convexity of $\|\cdot\|^2$,

$$\begin{aligned} \left\| \frac{\xi_m + \xi_n}{2} - x \right\|^2 &\leq \frac{1}{2} (\|\xi_m - x\|^2 + \|\xi_n - x\|^2), \\ \left\| L \left(\frac{\xi_m + \xi_n}{2} \right) \right\|^2 &= \frac{1}{4} \|L\xi_m + L\xi_n\|^2 \\ &\leq \frac{1}{2} (\|L\xi_m\|^2 + \|L\xi_n\|^2). \end{aligned}$$

It follows that

$$\limsup_{m, n \rightarrow \infty} G \left(\frac{\xi_m + \xi_n}{2} \right) \leq \frac{1}{2} \left(\lim_{m \rightarrow \infty} G(\xi_m) + \lim_{n \rightarrow \infty} G(\xi_n) \right) = \sigma^2.$$

Hence,

$$\lim_{m, n \rightarrow \infty} G \left(\frac{\xi_m + \xi_n}{2} \right) = \sigma^2.$$

The parallelogram identity

$$2\|f\|^2 + 2\|g\|^2 = \|f + g\|^2 + \|f - g\|^2 \quad (51)$$

yields

$$\begin{aligned} &2(G(\xi_m) + G(\xi_n)) \\ &= 2(\|\xi_m - x\|^2 + \|L\xi_m\|^2) + 2(\|\xi_n - x\|^2 + \|L\xi_n\|^2) \\ &= \|\xi_m + \xi_n - 2x\|^2 + \|\xi_m - \xi_n\|^2 + \|L(\xi_m + \xi_n)\|^2 \\ &\quad + \|L(\xi_m - \xi_n)\|^2. \end{aligned}$$

Therefore, as $m, n \rightarrow \infty$, we have

$$\begin{aligned} & \|\xi_m - \xi_n\|^2 + \|L(\xi_m - \xi_n)\|^2 \\ &= 2(G(\xi_m) + G(\xi_n)) - 4G\left(\frac{\xi_m + \xi_n}{2}\right) \rightarrow 0. \end{aligned}$$

It follows that

$$\lim_{m, n \rightarrow \infty} \|\xi_m - \xi_n\| = 0, \quad \lim_{m, n \rightarrow \infty} \|L\xi_m - L\xi_n\| = 0.$$

Therefore, the sequence ξ_n converges in norm to an element $\xi \in H_0$, and the sequence $L\xi_n$ converges in norm to an element $g \in H$. By the definition of the closed operator we have $\xi \in \mathcal{D}(L)$ and $g = L\xi$. Moreover, we have

$$\lim_{n \rightarrow \infty} \|L\xi_n\| = \|L\xi\|.$$

Finally, we obtain

$$\sigma^2 = \lim_{n \rightarrow \infty} G(\xi_n) = G(\xi), \quad \xi \in H_0 \cap \mathcal{D}(L),$$

as required. \square

Proof of Proposition 2.2: Assume that ξ_1 and ξ_2 are two distinct solutions of the problem (11), i.e.

$$G(\xi_1) = G(\xi_2) = \sigma^2.$$

By using the parallelogram identity (51), we have

$$\begin{aligned} & G\left(\frac{\xi_1 + \xi_2}{2}\right) \\ &= \frac{1}{2}(G(\xi_1) + G(\xi_2)) - \left\|\frac{\xi_1 - \xi_2}{2}\right\|^2 - \left\|\frac{L\xi_1 - L\xi_2}{2}\right\|^2 \\ &\leq \sigma^2 - \frac{1}{4} \|\xi_1 - \xi_2\|^2 < \sigma^2, \end{aligned}$$

which is impossible. The contradiction proves the proposition. \square

Proof of Proposition 2.4: Let L^* be the operator adjoint to L . Since L is a closed operator with the dense domain, L^*L is a self-adjoint non-negative operator, cf. [2, Chapter IV]. Therefore, $(I + L^*L)^{-1}$ is a bounded self-adjoint operator. By letting $A := I + L^*L$, we have

$$\begin{aligned} & \|y - x\|^2 + \|Ly\|^2 = \|x\|^2 - (x, y) - (y, x) + (Ay, y) \\ &= \|x\|^2 + \|A^{1/2}y - A^{-1/2}x\|^2 - \|A^{-1/2}x\|^2 \\ &\geq \|x\|^2 - \|A^{-1/2}x\|^2, \end{aligned}$$

and the equality is attained iff $y = \xi = A^{-1}x$. \square

Proof of Proposition 2.5: Let ξ be a solution of Problem (11). Then for every $Y \in H_0 \cap \mathcal{D}(L)$ we have

$$m := \|\xi - x\|^2 + \|L\xi\|^2 \leq \|Y - x\|^2 + \|LY\|^2.$$

Fix an arbitrary $h \in H_0 \cap \mathcal{D}(L)$; then $\xi + \varepsilon h \in H_0 \cap \mathcal{D}(L)$ for all real ε . We have

$$\begin{aligned} m &\leq \|\xi + \varepsilon h - x\|^2 + \|L(\xi + \varepsilon h)\|^2 \\ &= m + 2\varepsilon [\Re(\xi - x, h) + \Re(L\xi, Lh)] \\ &\quad + \varepsilon^2 [\|h\|^2 + \|Lh\|^2]. \end{aligned}$$

It follows that

$$\Re(\xi - x, h) + \Re(L\xi, Lh) = 0.$$

By replacing ε with $i\varepsilon$ we obtain

$$\Im(\xi - x, h) + \Im(L\xi, Lh) = 0.$$

By adding up two equalities we arrive at (12).

We establish now the uniqueness of the solution for (12). Let $\xi_1, \xi_2 \in H_0 \cap \mathcal{D}(L)$ satisfy equalities

$$\begin{aligned} (\xi_1 - x, h) + (L\xi_1, Lh) &= 0, \\ (\xi_2 - x, h) + (L\xi_2, Lh) &= 0, \end{aligned}$$

for all $h \in H_0 \cap \mathcal{D}(L)$. It follows that

$$(\xi_1 - \xi_2, h) + (L(\xi_1 - \xi_2), Lh) = 0.$$

By plugging $h = \xi_1 - \xi_2$ into this equality, we arrive at

$$\|\xi_1 - \xi_2\|^2 + \|L(\xi_1 - \xi_2)\|^2 = 0,$$

hence, $\xi_1 = \xi_2$. \square

Acknowledgements

M.A. Lifshits work was supported by RFBR grant 16-01-00258.

References

- [1] Akhiezer, N.I. Lectures on Approximation Theory. Nauka, 1965.
- [2] Akhiezer, N.I., Glazman, I.M. Theory of Linear Operators in Hilbert Space. Dover, 1993.
- [3] Bernstein, S.N. Sur une propriété de fonctions entières. Comptes Rendus Acad. Sci. Paris, t.176, 1603–1605 (1923).
- [4] Doob, J.L. Stochastic Processes. Wiley, N.Y. (1990).
- [5] Doob, J.L. Time series and harmonic analysis. In: Proc. Berkeley Sympos. Math. Statist. Probab., 303–344, Berkeley–Los Angeles, University of California Press, 1949.
- [6] Grenander, U., Szegö, G. Toeplitz Forms and Their Applications. AMS Chelsea Publ. (1998).
- [7] Hoffman, K. Banach Spaces of Analytic Functions, Prentice-Hall, Englewood Cliffs, N.J. (1962).
- [8] Kaj, I., Taqqu, M. S. Convergence to fractional Brownian motion and to the Telecom process: the integral representation approach. In: In and Out of Equilibrium 2. Ed. V. Sidoravicius and M. E. Vares. Ser: Progress in Probability **60**, Basel: Birkhäuser Verlag, 2008, pp. 383–427.
- [9] Kabluchko, Z., Lifshits M.A. Least energy approximation for processes with stationary increments. Preprint [arxiv 1506.08369](https://arxiv.org/abs/1506.08369). To appear in J. Theor. Probab.
- [10] Kolmogorov A.N. Interpolation and extrapolation of stationary random sequences. In: Selected Works of A.N. Kolmogorov, v.II, Probability and mathematical Statistics. Kluwer, Dordrecht, 1992. Original Russian edition: Izvestia AN SSSR. Ser. Math., v.5, pp. 3-14, 1941.
- [11] Lifshits, M. Random Processes by Example, World Scientific, Singapore, 2015.
- [12] Paley, R.E.A.C., Wiener, N. Fourier Transformations in the Complex Domain, AMS, Providence, 1934.
- [13] Rozanov Yu.A. Stationary Random Processes, Holden–Day, 1967. Original Russian edition: Nauka, Moscow, 1963.
- [14] Rudin, W. Real and Complex Analysis. Third Edition. Mc.Graw–Hill, 1987.
- [15] Yaglom, A.M. An Introduction to the Theory of Stationary Random Functions. Revised English edition. Dover, New York, 2004.
- [16] Wiener, N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Technology Press of MIT, Cambridge, 1949.